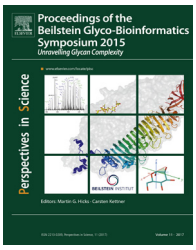
Perspectives in Science (2017) 11, 40—44

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/pisc



Bioinformatics of glycosaminoglycans☆

Han Hu[a,b], Yang Mao[b,c], Yu Huang[b,c], Cheng Lin[b,c], Joseph Zaia[a,b,c,*]

[a] *Bioinformatics Program, Boston University, Boston, MA, USA*
[b] *Center for Biomedical Mass Spectrometry, Boston University, Boston, MA, USA*
[c] *Department of Biochemistry, Boston University, Boston, MA, USA*

Received 2 November 2015; accepted 22 January 2016
Available online 20 October 2016

KEYWORDS
Glycosaminoglycan;
Heparan sulfate;
Tandem mass spectrometry;
Electron transfer dissociation;
Divide-and-conquer

Summary Cell surface heparan sulfates modulate many signalling pathways by binding growth factors and growth factor receptors. Expressed in a spatially and temporally regulated manner, these highly sulfated polysaccharides play important roles in all aspects of animal physiology. To understand heparan sulfate-protein binding, it is necessary to develop instrumental sequencing methods. Towards this end, we and others have demonstrated the effectiveness of activated electron dissociation (ExD) tandem mass spectrometry. The value in the ExD approach is that extremely rich tandem mass spectra are produced. The challenge is that bioinformatics methods are needed to convert the raw data into HS saccharide sequences. In this article we describe HS—SEQ, an algorithm developed for this purpose.

Introduction

Heparan sulfate (HS) is a linear sulfated polysaccharide that consists of repeating uronic acid-hexosamine disaccharide units. Biosynthesized in the endoplasmic reticulum and Golgi apparatus as a nascent co-polymer, HS chains are acted upon by a series of biosynthetic enzymes (Bulow and Hobert, 2006) as shown in Fig. 1. The HS domain structure results from the actions of *N*-deactylase/*N*-sulfotransferase enzymes that remove selected glucosamine *N*-acetate groups and replace them with *N*-sulfate groups. A glucuronic acid C5 epimerase then coverts a subset of uronic acid residues to iduronic acid. A series of *O*-sulfotransferases then adds sulfate groups to specific sites on the monosaccharide residues see Fig. 2. These enzymes add sulfate to targets primarily in the *N*-sulfated domains. The resulting mature chains have domains of high sulfation interspersed with those with low sulfation and those with intermediate sulfation.

* Corresponding author at: Dept. of Biochemistry, Boston University, Boston, MA, USA.
E-mail address: jzaia@bu.edu (J. Zaia).

http://dx.doi.org/10.1016/j.pisc.2016.01.014

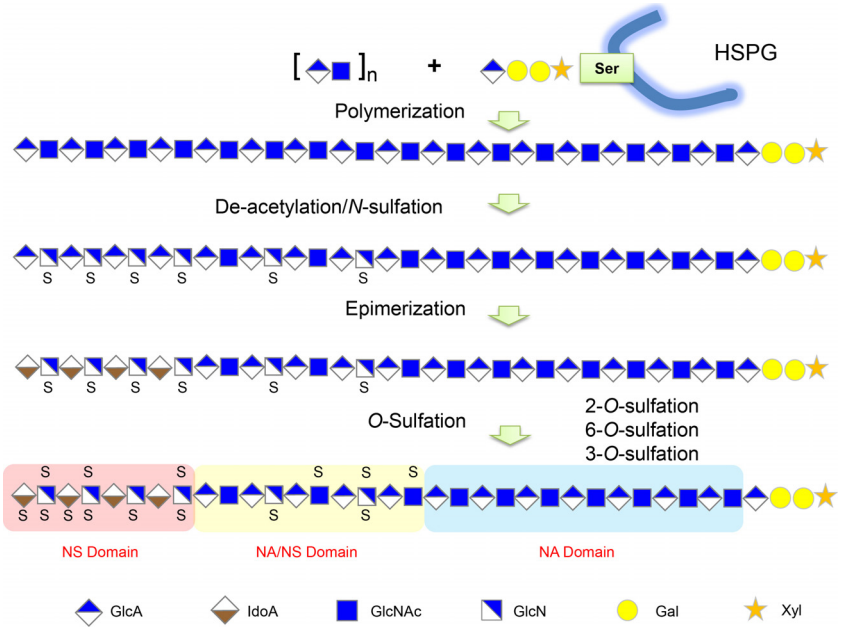


Figure 1 Biosynthesis of heparan sulfate.

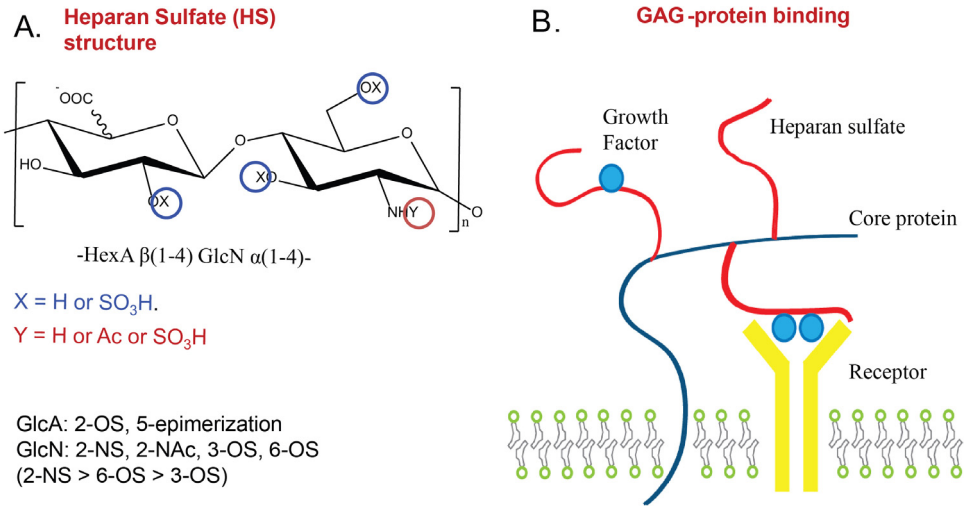


Figure 2 Overview of heparan sulfate structure. (A) Repeating disaccharide units with positions of sulfation indicated; (B) illustration of binding among HS chains growth factors and growth factor receptors.

Heparan sulfate is found in metazoans (DeAngelis, 2002) and is required for embryogenesis, development (Perrimon and Bernfield, 2000), and normal functioning of all physiological systems (Bishop et al., 2007). Heparin is a highly sulfated form of HS found in granulated hematopoietic lineage cells. Heparin derived from porcine intestinal mucosa mast cells is used as an anticoagulant in hospital surgical suites worldwide. Low molecular weight heparins (LMWH), including enoxaparin and dalteparin, are produced by limited depolymerization of heparin and used to treat deep vein thrombosis and other blood clotting disorders (Samama, 1992). The recent approval of biogeneric forms of LMWH reflected the power of analytical methods in characterizing these extremely complex biopolymeric mixtures so as to demonstrate acceptably low residual risk in the reverse engineered products.

Heparan sulfate is found on the surfaces of most animal cell types where it binds to many families of growth factors and growth factor receptors (Bernfield et al., 1999). The structure of HS expression varies among cell types and as a function of developmental and disease states. In normal tissue, cells receive signals from their surroundings that depend on interactions among cell surface HS, growth factors and growth factor receptors. In pathophysiological states including cancers, cellular responses to growth factor signals become dysregulated. These dysregulated states are often associated with changes in the structures of HS chains and the signaling proteins to which they bind. Thus,